

make + knitr

- make
 - description of dependencies among files, commands to (re)generate them
 - recalculates based on modification time of files
 - parallelizable
- knitr
 - interspersed code and explanatory text
 - dependencies based on text of code blocks
 - does not recompute based on changes of data files

make

greylist.bed: input_untreat.bam

Rscript generate_greylist.R input_untreat.bam greylist.bed

sample1_untreat_filt.bam: sample1_untreat.bam greylist.bed

filterByRegion greylist.bed sample1_untreat.bam sample1_untreat_filt.bam

sample1_untreat_peaks.xls: sample1_untreat_filt.bam input_untreat.bam

macs2 callpeak -t sample1_untreat_filt.bam -c input_untreat.bam \
--name sample_untreat_peaks

analysis.html: analysis.Rmd sample1_untreat_peaks.xls sample2_untreat_peaks.xls

Rscript -e rmarkdown::render(analysis.Rmd)

%.o: %.c

gcc -c \$< -o

knitr

```
---  
title: "Motif Analysis: FOXA1, GRHL2, H3K4me2"  
author: "Gord Brown"  
date: "`r Sys.Date()`"  
output: html_document  
---
```

```
```${r initialize}  
library(rtracklayer)
library(GenomicRanges)
interval <- 100
sample <- 1000
source_dir <- "bed"
sampled_peaks_dir <- "sampled"
```
```

Overall, the process is to sample ``r sample`` peaks from the upper half of each bed file, then generate bed files with intervals ``r interval`` bp wide, centred on the centres of the supplied bed files.

If the bed file is less than ``r sample*2`` intervals, but more than ``r sample``, then the ``r sample`` samples are from the whole file; if less than ``r sample`` intervals, the whole set is used.

...

Motif Analysis: FOXA1, GRHL2, H3K4me2

Gord Brown

2015-11-26

```
library(rtracklayer)
```

```
## Loading required package: methods
```

...

```
library(GenomicRanges)
interval <- 100
sample <- 1000
source_dir <- "bed"
sampled_peaks_dir <- "sampled"
```

Overall, the process is to sample 1000 peaks from the upper half of each bed file, then generate bed files with intervals 100 bp wide, centred on the centres of the supplied bed files.

If the bed file is less than 2000 intervals, but more than 1000, then the 1000 samples are from the whole file; if less than 1000 intervals, then a random set is used.